



# Pipelines de ETL e Machine Learning com Apache Spark

Ambiente de Desenvolvimento do Spark  
Scripts, Notebooks e Spark Shell

O Apache Spark oferece várias ferramentas e interfaces para facilitar o desenvolvimento, teste e depuração de aplicações Spark. Vamos discutir os três componentes principais: Scripts, Notebooks e Spark Shell.

- **Scripts:**

Os scripts Spark são basicamente programas escritos em uma linguagem de programação suportada pelo Spark (como Scala, Python ou Java) e podem ser executados no cluster Spark.

São ideais para aplicações de produção e tarefas que precisam ser agendadas ou automatizadas.

Os scripts podem ser empacotados como JARs (para Scala e Java) ou como arquivos Python e depois submetidos ao cluster usando o `spark-submit`.

- **Notebooks:**

Notebooks, como encontrados com Jupyter ou Databricks, oferecem uma interface interativa para desenvolvimento e análise de dados com Spark via interface web.

Permite escrever código, visualizar resultados e criar visualizações em um único ambiente.

São especialmente populares entre Cientistas de Dados para análise exploratória de dados, pois facilitam a visualização e o compartilhamento de insights.

O código nos notebooks pode ser exportado e transformado em scripts para produção.

- **Spark Shell:**

O Spark Shell é um REPL (Read-Evaluate-Print Loop) interativo para Scala (e PySpark para Python) que permite aos usuários executar comandos Spark em tempo real.

É uma ferramenta excelente para aprendizado, prototipagem rápida e depuração.

Os usuários podem iniciar o Spark Shell diretamente do terminal e começar a interagir com o cluster Spark imediatamente.

Comandos e operações executados no Spark Shell podem ser facilmente transferidos para scripts ou notebooks para desenvolvimento adicional.

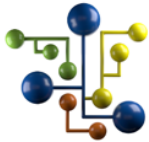
## ***Pipelines de ETL e Machine Learning com Apache Spark***

---

O ambiente de desenvolvimento do Spark é versátil e atende a diferentes necessidades, desde a análise exploratória de dados em notebooks até o desenvolvimento de aplicações de produção em scripts.

O Spark Shell serve como uma ponte entre esses dois extremos, oferecendo uma plataforma interativa para teste e depuração.

Neste curso usaremos principalmente a primeira opção, os scripts.



**Equipe DSA**

Muito Obrigado!  
Continue Trilhando Uma Excelente Jornada de Aprendizagem.